**Project Report: Development of IOCL ER Department Application**

**Project Overview:**

The project aimed to develop an application for the ER department of IOCL, enabling them to upload various files, extract data from these files, and interact with a chatbot to retrieve relevant information. The primary focus was on PDF documents, with provisions for other formats like TXT, CSV, DOC, and DOCX.

**Project Duration:**

**March 11th, 2024, to March 20th, 2024**

Key Tasks Accomplished:

**Environment Setup and Library Integration:**

Integrated necessary libraries including Streamlit, PyPDF2, and dotenv.

Utilized Streamlit for developing the user interface and interactive elements.

**File Upload and Text Extraction:**

Implemented functionalities to allow users to upload multiple PDF documents.

Developed methods to extract text from uploaded documents using PyPDF2 library for PDF files and text decoding for other formats.

**Text Chunking:**

Employed a text splitter module to divide the extracted text into manageable chunks for efficient processing.

Text was chunked to enhance the performance of subsequent operations, such as vectorization and retrieval.

**Vectorization and Retrieval:**

Integrated OpenAIEmbeddings to generate embeddings for text chunks.

Utilized FAISS to create a vector store from the text chunks, facilitating efficient retrieval of relevant information.

**Chatbot Implementation:**

Integrated a chatbot using ChatOpenAI to enable users to interact with the application.

Implemented conversational memory to maintain context and improve the chatbot's responsiveness.

User Interface Design:

**Designed an intuitive user interface using Streamlit.**

Provided features for users to input questions about the uploaded documents and receive responses from the chatbot.

**Testing and Deployment:**

Conducted thorough testing to ensure the functionality and reliability of the application.

Deployed the application for internal testing and feedback collection.